Best Practice Poster:
MARC to schema.org: Providing Better Access to UIUC Library Holdings Data

Timothy Cole
University of Illinois at Urbana-Champaign, United States
t-cole3@illinois.edu

Michael Norman
University of Illinois at Urbana-Champaign, United States
manorman@illinois.edu

Patricia Lampron
University of Illinois at Urbana-Champaign, United States
lampron2@illinois.edu

William Weathers
University of Illinois at Urbana-Champaign, United States
weathrs2@illinois.edu

Ayla Stein
University of Illinois at Urbana-Champaign, United States
astein@illinois.edu

M. Janina Sarol
University of Illinois at Urbana-Champaign, United States
mjsarol@illinois.edu

Myung-Ja Han
University of Illinois at Urbana-Champaign, United States
mhan3@illinois.edu

Keywords: MARC, schema.org; MARCXML; bibliographic description; MODS; holdings data.

1. Introduction

Taking advantage of the Web as a means for disseminating large datasets, libraries have begun publishing their bibliographic metadata on the Web—e.g., the University of Michigan, the University of Florida, and Harvard University. Initially, most libraries focused on releasing their catalogs as MARCXML, however, MARC consists primarily of string data with few, if any, URIs linking to ontologies or related resources. MARCXML was not designed for use with RDF. Libraries are now experimenting with disseminating catalogs as linked open data in other serializations, e.g., OCLC, and the British Library. Semantics compatible with RDF are being used, but specific schemes vary. Detail about holdings associated with bibliographic descriptions is still lacking, e.g., the volumes of a described serial title held by the library are not enumerated. This last seems a significant omission given that libraries are uniquely positioned to provide this information. The University of Illinois at Urbana-Champaign (UIUC) Library has released 5.5 million bibliographic catalog records that include detailed local holdings information to allow consumers to know exactly which volumes or parts of the creative work described are available at UIUC. MARCXML serializations are available for downloading now. MODS serializations enriched with links to name and subject authorities and RDF serializations (using schema.org semantics) will soon be available. This poster reports on the development of workflows for this project, on the multiple formats of catalog metadata being made available through these workflows, and on the lessons learned to date.

1 http://www.lib.umich.edu/library-information-technology/open-access-bibliographic-records-available-download-and-use
2 http://www.uflib.ufl.edu/catmet/creativecommons.html
3 http://openmetadata.lib.harvard.edu/bibdata
4 http://www.worldcat.org/
5 http://bnb.data.bl.uk/
2. MARCXML with physical holdings information

As a first step, we created MARCXML bibliographic descriptions for each physical volume the library holds with selected volume-specific information (e.g., barcode) recorded in the 955 local data field. With a simple VB.NET program, we collapsed volume-level records associated with a single bibliographic entity into one bibliographic record that contains all holding and item level information for associated volumes and parts in repeated MARC 852 data fields as shown in Figure 1.

```
<marc:datafield tag="852" ind1="0" ind2=" ">
  <marc:subfield code="a">IU</marc:subfield>
  <marc:subfield code="b">Rare Book &amp; Manuscript Library [non-circulating]</marc:subfield>
  <marc:subfield code="h">099</marc:subfield> <!-- classification number -->
  <marc:subfield code="i">Ab3</marc:subfield> <!-- cutter -->
  <marc:subfield code="p">30112066264109</marc:subfield> <!-- barcode -->
  <marc:subfield code="t">1</marc:subfield> <!-- copy number -->
</marc:datafield>
```

FIG 1: Example of MARC XML 852 data field used to record physical holdings

3. MODS Transformation & Adding Links

The transformation of MARCXML with holdings information in 852 data fields into MODS is based on the Library of Congress (LC) MARC to MODS recommendations.6 (We differ slightly from the LC mapping recommendations in how we treat enumeration/chronology, copy number, and barcode.) Each 852 data field is mapped to a MODS <location> element. 852 subfield a is mapped to <location> sub-element <physicalLocation>; all other 852 subfields map to sub-elements of a single <copyInformation> element, within the <holdingSimple> subelement of <location>. Figure 2 displays the 852 data field of Figure 1 transformed to MODS.

```
<mods:location>
  <physicalLocation displayLabel="Institution Code">IU</physicalLocation>
  <holdingSimple>
    <copyInformation>
      <subLocation> Rare Book &amp; Manuscript Library [non-circulating]</subLocation>
      <shelfLocator>099 Ab3</shelfLocator>
      <note displayLabel="Copy Number">1</note>
      <note displayLabel="Barcode">30112066264109</note>
    </copyInformation>
  </holdingSimple>
</mods:location>
```

FIG 2: MARC 852 data field transformed to MODS

After transforming MARCXML records to MODS, a Python script is invoked to search VIAF for URIs matching values in the MODS <name> element, as transformed from MARCXML data fields 100, 110, 111, 700, 710, 711, and 720. When found, URIs are added to the MODS <name> element replacing the string values. When searching VIAF, we use complete name information, birth date, and death date (as available). Only exact matches in VIAF are recorded. The same script searches LCSH Linked Data Services7 to find subject heading URIs, which are then also

---

6 http://www.loc.gov/standards/mods/userguide/location.html
7 http://id.loc.gov/
added to the MODS <subject> element. If no match is found, the text string remains as the value for the field.

4. Transformation to RDF and schema.org

The MODS metadata enriched with links to name and subject authorities are transformed into schema.org semantics. These are disseminated one-by-one as RDFa (within HTML styled for presentation to end-users), via bulk downloading (as RDF/XML or JSON-LD), and via a SPARQL endpoint. Transformation of bibliographic metadata from MODS to schema.org is straightforward (though arguably the distinction between work and manifestation is further blurred). However, transforming holdings to schema.org is challenging. Based on earlier experimentation at OCLC and our interpretation of relevant W3C Schema Bib Extend Community Group guidelines,8 we mapped each holding as a schema.org <offer> entity.

Conclusion

The goal of this poster is two-fold:

- sharing with the community practices and workflow implementations developed at UIUC for disseminating traditional library data in multiple formats and serializations; and,
- gaining feedback on the mapping and modeling decisions made in transforming detailed MARC bibliographic and holdings data into linked open data.

8 http://www.w3.org/community/schemabibex/wiki/Holdings_via_Offer